ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Molecular Dynamics Simulations: Computational Strategies to Delineate Intra and Inter Molecular Interactions.

Biswajit Gorai, Mahaboob Raashma and Thirunavukkarasu Sivaraman*.

Structural Biology Lab, Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA University, Thanjavur - 613 401, Tamil Nadu, India.

ABSTRACT

Molecular dynamics simulations play pivotal roles on unraveling residues that are essential for structural integrities and functional activities of proteins and as well on analyzing interfaces of proteins-ligands/macromolecules complexes at atomistic resolution in near physiological conditions. In this review article, we have systematically discussed salient features and limitations of different types of MD simulations (all-atom MD, coarse-grained MD and steered MD) and various parameters (force fields, integrating algorithms, simulation models and implicit/explicit water systems) that are being used in the simulations studies in a concise and forthright manner.

Keywords: All-atom MD; coarse-grained MD; steered MD; drug design

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Molecular dynamics (MD) simulations are unique strategies to deal with motions of molecules as a function of time at atomic detail [1, 2] and significances of the MD simulations on understanding various complex areas of biosciences (Anfinsen's hypothesis and Levinthal paradox of protein folding; aggregation and thermodynamic stabilities of proteins; mechanisms of molecular interactions; membrane dynamics, drug designing and so on) have been unequivocally demonstrated in the literature [3-6]. The MD simulation was first applied on an assembly of hard-spheres by Alder and Wainwright [7, 8] and subsequently, Rahman demonstrated MD simulations of liquid argon at 94.4 K [9] and as well liquid water using potential functions [10]. The major breakthrough in the MD studies was brought by McCammon and coworkers by successfully demonstrating MD simulation of bovine pancreatic trypsin inhibitor (BPTI) in vacuum for 9.2 ps at 295 K [11]. MD simulations have now become eminent computational strategies to study larger biological systems and their complex processes by taking advantages of high-performance computational facilities and highly sophisticated MD algorithms available to the scientific community [12].

In the time span of nearly half-century (1977-2015), quite a large number of research papers (> 30000) have been published on MD simulations of bio-macromolecules and several review articles on MD have also been published by various research groups from all around the world [13]. In this review article, we have herein discussed the unique merits, limitations and scopes of all-atom MD, coarse-grained MD and steered MD simulations in a clear, concise and comprehensive manner. Moreover, limitations of different models accounting explicit and as well implicit water molecules have also been discussed. Thus, we trust that the review would be useful to choose pertinent simulations strategies in order to carry-out MD simulations of bio-macromolecules in an efficient manner.

MD simulations

Position of each atom in a molecule can be calculated at every step of MD simulations using integrating algorithms designed on the basis of Newton's laws of motion [14, 15]. All integrating algorithms consider previous atomic position, velocity and acceleration of an atom in order to determine its new positions approximated by a Taylor series [12]. Integrating algorithms such as Verlet [16], leap-frog [17] and Beeman [18] are being generally employed in MD simulations. For instance, positions of atoms in a simulated molecule can be calculated after a short and finite interval of 'Δt' from a specific simulation time 't' as shown herein:

$$x(t+\Delta t)=2x(t)-x(t-\Delta t)+\frac{d^2x(t)}{dt^2}\Delta t^2 \qquad (1)$$

Where in, 'x' denotes distance between old and new positions of respective atoms. And, every step in MD simulations consists of computationally expensive calculations of bonded and non-bonded interactions for each atom in molecules. Bonded parameters such as bond distances, bond angles and dihedral angles that are represented on the basis of Hooke's law and non-bonding interactions accounting non-polar attractions (Lenard-Jones energies) and polar attractions (Coulomb energies) among atoms of molecules can be calculated using potential energy functions derived from ab initio quantum mechanical calculations and experimental studies (especially spectroscopy methods) of small molecules [19] as shown in equation 2.

$$E_{total}=\underbrace{\sum_{bonds}K_r(r-r_{eq})^2+\sum_{angles}K_\theta(\theta-\theta_{eq})^2+\sum_{dihedrals}\frac{V_n}{2}\left[1+\cos(n\phi-\gamma)\right]}_{\text{Bonded interactions}}+\underbrace{\sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}}-\frac{B_{ij}}{R_{ij}^{6}}+\frac{q_iq_j}{\varepsilon R_{ij}}\right]}_{\text{Non-bonded interactions}} \qquad (2)$$

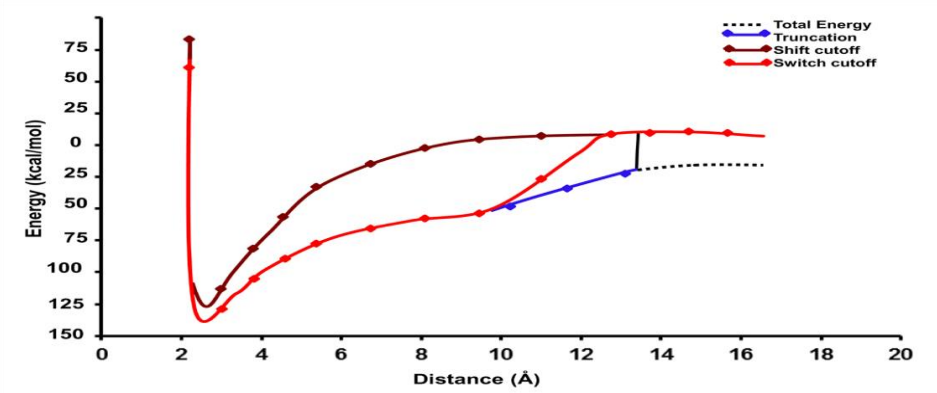
wherein, K_r, K_θ and V_n represent force constants for bond length, bond angle and dihedral angle, respectively; r, θ and ϕ denote the bond length, bond angle and dihedral angle, respectively; r_{eq}, θ_{eq} and γ are respective equilibrium values for each bonded parameters mentioned above. In case of non-bonded interactions, $A = 4\varepsilon\sigma^{12}$ and $B = 4\varepsilon\sigma^{6}$ (ε is the depth of potential signifying attractive force between two particles and σ is the

ISSN: 0975-8585

distance at which the intermolecular potential between the two particles is zero); R_{ij} is the distance of separation of two atoms 'i' and 'j' with 'q_i' and 'q_j' partial charges, respectively; ε is dielectric constant of medium used in MD simulations.

All-atom MD simulations: The all-atom MD simulations are generally used to obtain structural details at atomistic resolution at a large expense of computational power. The first all-atom simulation of a globular protein, bovine pancreatic trypsin inhibitor (BPTI), in a truncated octahedron box with 1462 water molecules for 20 ps was performed by Gunsteren and Berendsen [20]. Force fields such as AMBER [21], CHARMM [22] and OPLS-AA [23] are particularly being used for these types of MD simulations. Non-bonded interactions of trajectory structures derived from the MD simulations can be computed by three different methods: truncation, shift and switch [24]. In truncation method, the interactions among the atoms of molecules are set as zero beyond cut-off distance defined for calculations. In shift method, the whole potential energy surface will be altered to omit an interaction after cut-off distance. In switch method, two cut-off distances are used. The potential follows the usual energy function within first cut-off and later turn to zero between first and last cut-off and the 'switch scheme' is generally recommended for long range cut-off potentials. Functions of all the three methods are schematically represented in Figure 1.Though introduction of cut-off functions help to speed-up the calculations, some undesirable artifacts may be found in certain cases. To tackle those scenarios, programs such as Ewald lattice-sum [25], Fast Multipole Method (FMM) [26], Reaction Field approximations [27], Extended Electrostatics model [28] can be effectively used to account crucial long-range electrostatic interactions present in the molecules generated in each frame of simulations.

Figure 1: Schematic representations of various methods that can be used to calculate non-bonding interactions of molecules in all-atom MD simulations.



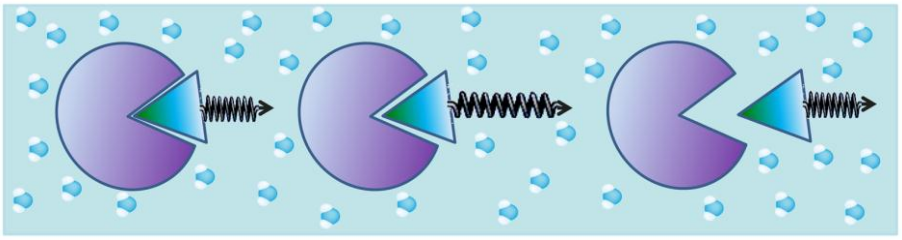
Coarse-grained MD (CGMD) simulations: Though all-atom MD simulations can be used to derive structural details at atomistic level resolution, the methods, (notwithstanding the sophistication of MD algorithms and computational facilities available to date) may not be useful to study larger macromolecular systems and as well biological processes occurring in microseconds or slower than the timescale. Coarse-grained MD simulations (CGMD) are excellent alternatives to the all-atom MD methods to efficiently tackle the time constraint barrier at the expense of atomic resolution. In the MD strategies, several atoms will be grouped together and represented as a 'bead'. Less number of beads means less computational cost for simulation [29]. Interestingly, the decrease of effective number of particles may increase integration time-step (even up to 50 fs) and consequently, the calculation can be speed-up to a factor of 10^6 or so vis-a-vis all-atom MD simulations. United-atom force fields such as Gromos [30] and OPLS [31] are appropriate choices for the CGMD simulations. Moreover, CG models available to date can be divided into two broad groups: structure-based and structure-independent models [32] and the former and later models use the physicochemical properties derived from target molecular structures and ab initio calculations, respectively. Of the two models, the former models are widely being used. Elastic network model [33], Gō models [34], MARTINI force field [35] and AMH (associated memory Hamiltonian) model [36] are the popular examples for the structure-based coarse-grained MD simulations. The CGMD simulations were successfully applied by Levitt and Warshel [37] to study events in the refolding of BPTI. In the simulations, each amino acid residue of the protein was represented in two sets, C_α and centroid of side chain defining position of the polypepdide chain and interaction sites, respectively. By this way, the degrees of freedom and number of interactions could be reduced by a factor of four and fifteen respectively and consequently computational costs were reduced by

many folds. Nowadays, the CGMD simulations are regularly being used as evidenced from literature to address protein folding, protein aggregation, protein-protein/nucleic acid interactions, membrane-protein dynamics, conformational sampling of multi-domain proteins, mechanism of bio-molecular motors (like F1-ATPase, myosin, kinesin etc.), ribosome, RNA polymerase and many more, obviously [38, 39, 32]. Notwithstanding the high speed of CGMD, inaccuracy on predicting atomic positions/motions and negating of non-native interactions are inherent limitations of the methods.

Steered MD (SMD) simulations: Steered molecular dynamics (SMD) is an enhanced sampling method, which applies external mechanical force to selected atoms of target molecules in order to mapping out specific conformational changes. In principle, SMD complements the experimental techniques such as atomic force microscopy (AFM), optical tweezers, biomembrane force probe and surface force apparatus experiments [40-44]. In SMD simulations, external force will be applied to pull an atom or a group in a defined direction and velocity using a harmonic spring and a diagrammatic representation of the phenomenon is depicted in Figure 2. The force required to pull the particles is calculated as shown in equation 3.

$$F = K (x_o + vt - x_t) \quad (3)$$

Figure 2: Illustration representing an external force applied to pull out a small molecule from its target molecule in steered molecular dynamics simulations.



In the above equation, 'x_o' and 'x_t' are the initial and final positions of particles, respectively, when they are pulled for time 't' with force constant 'K' and velocity 'v'. The SMD was first used to study the force required to dissociate the streptavidin-biotin complex [45] and the resultant computational data were presented to be in good agreement with data obtained from AFM of the complex. Since then, SMD simulations are regularly being used to study mechanism of detachment of inhibitor-protein complexes and as well stretching of several mechanical proteins (like immunoglobulin titin and fibronectin). Moreover, SMD simulations are emerging as a reliable tool in drug designing to characterize protein-ligand complexes from specificity point of views and the SMD simulations in conjunction with Jarzynski's [46] and weighted histogram analysis (WHAM) theorem [47] can be used to calculate potential of mean force (PMF) for receptor-inhibitor complex, which in turn can be effectively used to explore active and inactive ligands to target molecules. However, SMD can't be used to screen large small molecules database as molecular docking algorithms perform to achieve high throughput virtual screening of small chemical molecules. On the other hand, the SMD can be reliably used in a unique manner to calculate relative binding affinities of protein-ligand complexes and as well to examine epitope mapping of small chemical molecular ligands bound with protein targets [41, 48].

Water models for the MD simulations

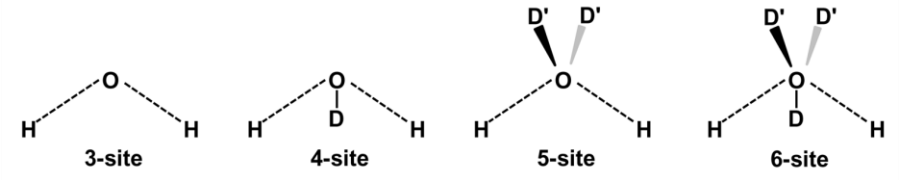
Water plays vital roles in almost every cellular reaction and hence water is an elixir of all living things of our planet. Amazing properties of water molecules such as high dielectric constant, hydrogen-bond formations, latent heat, densities, sublimation, surface tension, viscosity and acid-base properties make the task of defining parameters to water to be extremely tough. In case of MD simulations, water can be represented either as explicit (solvent molecules in a discrete fashion) models or as implicit (solvent molecules in a continuum fashion) models and the various models representing water molecules have been discussed in the following sections.

Explicit water models: Nearly 50 models of water have been proposed till date to accurately reproduce various properties of water molecules and the models are classified on the basis of their interaction sites in

ISSN: 0975-8585

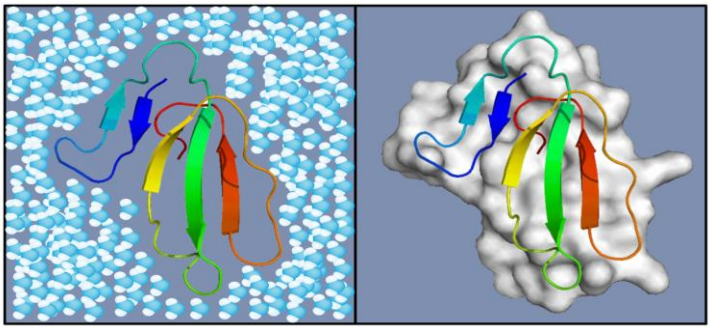
general (Figure 3). In 3-site water model, there are three interaction sites and the models are being extensively used in routine MD simulations due to their higher computer efficiency. Simple point charge (SPC) and transferable intermolecular potential (TIP3P) are the models belonging to this category [49, 50]. In 4-site water model, one dummy atom (partial charge with no mass) along the bisector of the HOH angle of water molecule was added to improve the electrostatic distribution around the molecule. Bernal-Fowler [51], TIPS2 [52], TIP4P [53], TIP4P-Ew [54], TIP4P/Ice [55], TIP4P/2005 [56], RWK2, SRWK [57] and KJ [58] are some of the models of this category. Water molecule is represented with two lone pairs in a tetrahedral geometry in 5-site water models and the models precisely reproduce experimental radial distribution function and density of water molecules. BNS, ST2, ST4 [59], TIP5P [60] and TIP5P-E [61] are popular 5-site water models. With three dummy atoms, 6-site water model was developed by Nada and van der Eerden [62] and the model precisely accounts melting point and as well density of ice comparing to other models.

Figure 3: Various models representing water molecule. Oxygen, Hydrogen and Dummy atoms are denoted as 'O', 'H' and 'D/D'', respectively.



Implicit water model: Implicit water environment in MD simulations is represented as continuum potential of mean force with average properties of water as determined from quantum mechanics and experimental methods [63-65]. Implicit solvent models can be classified into two major types: models based on accessible surface areas (ASA) and models based on electrostatic potentials [66, 67]. The electrostatic solvation free energy (solute-water interactions) can be calculated by using Poisson–Boltzmann (PB) theory [68] and the Generalized Born (GB) theory [69]. Several software packages such as APBS [70], CHARMM, Delphi [71], Jaguar [72], UHBD [73] and MEAD [74] are available to carry-out the free-energy calculation in an efficient manner. Figure 4 is a two-dimensional pictorial illustration for explicit and implicit water models of MD simulations. Recent computational advances facilitated implicit solvent models to successfully explore the areas of protein folding, structure refinements, membrane-protein interactions, free energies of proteins bound with small chemical molecules/peptides/proteins/nucleic acids [75-79]. Though 'implicit water models' substantially reduced dynamic timescale comparing to use of 'explicit water models' in MD simulations, the implicit water model posed inherent limitations on studying water-mediated interactions and dynamics of ions in ion channels, in particular [63, 80]. In other words, 'implicit water models' describe only non-specific interactions between water and solute molecules, whereas 'explicit water models' help to acquire specific interactions between water and solute molecules at high resolution. In conclusion, as MD simulations are emerging as reliable computational alternatives on divulging various biological processes that are posing many limitations to be characterized by experimental methods, we trust that the review will be very useful for computational and structural biologists to conduct exciting MD simulations on 'biomolecules' and 'drug designing' in near future.

Figure 4: A protein molecule is depicted in an explicit (left) and an implicit (right) solvent system. In the implicit water model, the white and blue shaded regions represent dielectric constant (ε) of 1 and 80, respectively. The protein and water molecules are represented in cartoon and sphere models, respectively.



ACKNOWLEDGEMENTS

This research work is partly supported by research grants'BIC/11(20)/2013'from Indian Council of Medical Research (ICMR) and 'BT/PR7822/BID/7/433/2013'from Department of Biotechnology (DBT), New Delhi, India.

REFERENCES

[1] Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Annu Rev Biophys 2012; 41: 429–52.
[2] Schlick T, Collepardo-Guevara R, Halvorsen LA, Jung S, Xiao X. Q Rev Biophys 2011; 44(2): 191-228.
[3] Karplus M, McCammon JA. Nat Struct Biol 2002; 9(9): 646-652.
[4] Durrant JD, McCammon JA. BMC Biol 2011, 9:71.
[5] Beck DAC, Daggett V. Methods 2004; 34: 112–120.
[6] Salsbury FR Jr. Curr Opin Pharmacol 2010; 10(6): 738–744.
[7] Alder BJ, Wainwright TEJ. J Chem Phys 1957; 27: 1208-1209.
[8] Alder BJ, Wainwright TEJ. J Chem Phys 1959; 31: 459-466.
[9] Rahman, A. Phys Rev A 1964; 136: 405-411.
[10] Stillinger FH, Rahman A. J Chem Phys 1974; 60: 1545-1557.
[11] McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature (Lond.) 1977; 267(5612): 585-590.
[12] Adcock SA, McCammon JA. Chem Rev 2006; 106(5): 1589–1615.
[13] Guliaev AB, Cheng S, Hang B. World J Methodol 2012; 2(6): 42-49.
[14] Van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glattli A, Hunenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt FA, Yu HB. Biomolecular Modeling: Goals, Problems, Perspectives. Angew ChemInt Ed 2006; 45: 4064-4092.
[15] Hashem Y, Auffinger P. Methods 2009; 47: 187-197.
[16] Swope WC, Andersen HC, Berens PH, Wilson KR. J Chem Phys 1982; 76: 637-649.
[17] Birdsall CK, Langdon AB. Effects of the Finite Time Step. In: Plasma Physics via Computer Simulation. Bristol, England, Taylor and Francis Group, 2005, pp 183-212.
[18] Beeman D. J Comput Phys 1976; 20(2): 130-139.
[19] Boas FE, Harbury PB. Curr Opin Struct Biol 2007; 17: 199-204.
[20] Van Gunsteren WF, Berendsen HJ. J Comput Aided Mol Des 1987; 1(2): 171-176
[21] Weiner PK, Kollman PA. J Comput Chem 1981; 2(3): 287-303.
[22] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S and Karplus M. J Comput Chem 1983; 4(2): 187-217.
[23] Jorgensen WL, Tirado-Rives J. J Am Chem Soc 1988;110(6): pp 1657–1666.
[24] Toxvaerd S, Dyre JC. J Chem Phys 2011; 134: 081102.
[25] Darden T, York D, Pedersen L. J Chem Phys 1993; 98: 10089-10092.
[26] Kurzak J, Pettitt BM. Mol Simul 2006; 32(10-11): 775-790.
[27] Lin Y, Baumketner A, Deng S, Xu Z, Jacobs D, Cai W. J Chem Phys 2009; 131; 154103.
[28] Stote RH, States DJ, Karplus M. J Chim Phys 1991; 88: 2419-2433.
[29] Tozzini V. Curr Opin Struct Biol 2005; 15:144–150.
[30] Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF. J Phys Chem A 1999; 103: 3596-3607.
[31] Jorgensen WL, Tirado-Rives J. J Am Chem Soc 1988; 110(6): 1657-1666.
[32] Takada S. Curr Opin Struct Biol 2012; 22: 130–137.
[33] Tirion MM. Phys Rev Lett 1996; 77(9): 1905-1908.
[34] Clementi C, Nymeyer H, Onuchic JN. J Mol Biol 2000; 298: 937-953.
[35] Marrink SJ, de Vries AH, Mark AE. J Phys Chem B 2004; 108: 750-760.
[36] Hardin C, Eastwood MP, Luthey-Schulten Z, Wolynes PG. PNAS 2000; 97(26): 14235-14240.
[37] Levitt M, Warshel A. Nature 1975; 253(5494): 694-698.
[38] Ha-Duong T. AdvExp Med Biol 2014; 805:157-169.
[39] Saunders MG, Voth GA. Annu Rev Biophys 2013; 42:73-93.
[40] Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. Structure 2009; 17(10): 1295-1306.
[41] Li MS, Mai BK. Curr Bioinform 2012; 7: 342-351.
[42] Isralewitz B, Baudry J, Gullingsrud J, Kosztin D, Schulten K. J Mol Graphics Modell 2001; 19: 13–25.
[43] Isralewitz B, Gao M, Schulten K. Curr Opin Struct Biol 2001; 11: 224-230.

[44] Lu H, Schulten K. Proteins Struct Funct Bioinf 1999; 35: 453-463.
[45] Grubmüller H, Heymann B, Tavan P. Science 1996; 271(5251): 997-999.
[46] Jarzynski, C. Phys Rev Lett 1997; 78(14): 2690-2693.
[47] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. J Comput Chem 1992; 13(8): 1011-1021.
[48] Bello M, Correa-Basurto J. Plos One 2013; 8(8): e72575.
[49] Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration.In: Intermolecular Forces. Pullman B, (eds). 14th ed. Dordrecht, Holland,D. Reidel Publishing Company, 1981, pp 331-342.
[50] Jorgensen WL. J Am Chem Soc 1981; 103(2): 335-340.
[51] Bernal JD, Fowler RH. J Chem Phys 1933; 1: 515.
[52] Jorgensen WL. J Chem Phys 1982; 77: 4156-4163.
[53] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J Chem Phys 1983; 79: 926-935.
[54] Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. J Chem Phys 2004; 120: 9665-9678.
[55] Abascal JLF, Sanz E, Fernández RG, Vega C. J Chem Phys 2005; 122: 234511.
[56] Abascal JLF, Vega C. J Chem Phys 2005;123: 234505.
[57] Reimers JR, Watts RO, Klein ML. Chem Phys 1982; 64: 95-114.
[58] Sprik M. J Chem Phys 1991; 95: 6762.
[59] Stillinger FH, Rahman AJ. Chem Phys 1974; 60: 1545-1557.
[60] Mahoney MW, Jorgensen WL. J Chem Phys 2000; 112: 8910-8922.
[61] Rick SW. J Chem Phys 2004; 120: 6085-6093.
[62] Nada H, van der Eerden JPJM. J Chem Phys 2003; 118: 7401-7413.
[63] Feig M, Brooks CL III. Curr Opin Struct Biol 2004; 14: 217-224.
[64] Chen J, Brooks CL III and Khandogin J. Curr Opin Struct Biol 2008; 18: 140-148.
[65] Chen J, Im W, Brooks CL III. J Am Chem Soc 2006; 128: 3728-3736.
[66] Ferrara P, Apostolakis J, Caflisch A. Proteins 2002; 46: 24-33.
[67] Hassan SA, Mehler EL. Proteins Struct Funct Bioinf 2002; 47(1): 45-61.
[68] Fogolari F, Brigo A, Molinari H. J Mol Recognit 2002; 15(6): 377-392.
[69] Bashford D, Case DA. Annu Rev Phys Chem 2000; 51: 129-152.
[70] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Proc Natl Acad Sci USA 2001; 98: 10037-10041.
[71] Rocchia W, Alexov E, Honig B. J Phys Chem B 2001; 105(28): 6507-6514.
[72] Cortis CM, Friesner RA. J Comput Chem 1997; 18: 1591-1608.
[73] Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin AV, Antosiewicz J, Gilson MK, Bagheri B, Scott LR, McCammon JA. Comput Phys Commun 1995; 91: 57-95.
[74] Bashford D. Scientific Computing in Object-Oriented Parallel Environments, Ishikawa Y, Oldehoeft RR, Reynders JVW, Tholburn M, Eds. An Object-oriented Programming Suite for Electrostatic Effects in Biological Molecules. Springer, Berlin, Germany, 1997, pp. 133–140.
[75] Ji C, Mei Y. Acc Chem Res 2014; 47(9): 2795-2803.
[76] Lazaridis T. Proteins 2005; 58(3): 518-527.
[77] Efremov RG, Volynsky PE, Nolde DE, Dubovskii PV, Arseniev AS. Biophys J 2002; 83(1): 144-53.
[78] Liu Y, Haddadian E, Sosnick TR, Freed KF, Gong H. Biophys J 2013; 105(5): 1248-57.
[79] Chocholoušová J, Feig M. J Phys Chem B 2006; 110 (34): 17240-17251.
[80] Kleinjung J, Fraternali F. Curr Opin Struct Biol 2014; 25(100): 126–134.